

# Increasing confidence for discerning species and population compositions from metabarcoding assays of environmental samples: case studies of fishes in the Laurentian Great Lakes and Wabash River

Matthew R. Snyder<sup>1,2,3</sup>, Carol A. Stepien<sup>2</sup>

1 *Genetics and Genomics Group, Department of Environmental Sciences, University of Toledo, (Lab group led by CAS relocated to 2 and 3 in 2016), Toledo, OH, 43606, USA*

2 *Genetics and Genomics Group, Pacific Marine Environmental Laboratory, National Oceanic and Atmospheric Administration, 7600 Sand Point Way NE, Seattle, WA, 98115, USA*

3 *Genetics and Genomics Group, Joint Institute for the Study of the Atmosphere and Ocean (JISAO), University of Washington, Seattle WA, 98195, USA*

Corresponding author: Carol A. Stepien ([carol.stepien@noaa.gov](mailto:carol.stepien@noaa.gov); [cstepien@uw.edu](mailto:cstepien@uw.edu))

**Academic editor:** Bernd Hänfling | **Received** 21 April 2020 | **Accepted** 11 August 2020 | **Published** 26 August 2020

## Abstract

Community composition data are essential for conservation management, facilitating identification of rare native and invasive species, along with abundant ones. However, traditional capture-based morphological surveys require considerable taxonomic expertise, are time consuming and expensive, can kill rare taxa and damage habitats, and often are prone to false negatives. Alternatively, metabarcoding assays can be used to assess the genetic identity and compositions of entire communities from environmental samples, comprising a more sensitive, less damaging, and relatively time- and cost-efficient approach. However, there is a trade-off between the stringency of bioinformatic filtering needed to remove false positives and the potential for false negatives. The present investigation thus evaluated use of four mitochondrial (mt) DNA metabarcoding assays and a customized bioinformatic pipeline to increase confidence in species identifications by removing false positives, while achieving high detection probability. Positive controls were used to calculate sequencing error, and results that fell below those cutoff values were removed, unless found with multiple assays. The performance of this approach was tested to discern and identify North American freshwater fishes using lab experiments (mock communities and aquarium experiments) and processing of a bulk ichthyoplankton sample. The method then was applied to field environmental (e) DNA water samples taken concomitant with electrofishing surveys and morphological identifications. This protocol detected 100% of species present in concomitant electrofishing surveys in the Wabash River and an additional 21 that were absent from traditional sampling. Using single 1 L water samples collected from just four locations, the metabarcoding assays discerned 73% of the total fish species that were discerned during four months of an extensive electrofishing river survey in the Maumee River, along with an additional nine species. In both rivers, total fish species diversity was best resolved when all four metabarcoding assays were used together, which identified 35 additional species missed by electrofishing. Ecological distinction and diversity levels among the fish communities also were better resolved with the metabarcoding assays than with morphological sampling and identifications, especially using all four assays together. At the population-level, metabarcoding analyses targeting the invasive round goby *Neogobius melanostomus* and the silver carp *Hypophthalmichthys molitrix* identified all population haplotype variants found using Sanger sequencing of morphologically sampled fish, along with additional intra-specific diversity, meriting further investigation. Overall findings demonstrated that the use of multiple metabarcoding assays and custom bioinformatics that filter potential error from true positive detections improves confidence in evaluating biodiversity.

## Key Words

community composition, cytochrome *b*, eDNA, Great Lakes, population variation, species detection, species diversity, 12S RNA



## Introduction

Assessments of species compositions and diversities of biological communities are fundamental for understanding their ecology (Elton 1966; Begon et al. 2006; Morin 2009), facilitating conservation efforts (Myers et al. 2000; Margules et al. 2002) and evaluating anthropogenic impacts (Attrill and Depledge 1997). Identification of rare and/or endangered species is of importance to fishery and conservation managers (Dobson et al. 1997; Margules et al. 2002), along with detection of non-native species (Allendorf and Lundquist 2003). However, capture-based surveys with morphological identifications are costly to conduct, require extensive taxonomic expertise, and are prone to false negatives (Attrill and Depledge 1997; Balmford and Gaston 1999; Darling and Mahon 2011).

Metabarcoding assays employing high-throughput sequencing (HTS) can be used for species identifications and calculations of community diversity, and are more sensitive, less damaging, and relatively time- and cost-efficient than are morphological determinations from capture-based surveys (Smart et al. 2016; Deiner et al. 2017). Some of the prior studies that have compared numbers and biomass determinations of captured taxa to the relative proportions of sequence reads returned from metabarcoding assays of water environmental (e) DNA samples found positive correlations (Hänfling et al. 2016; Thomsen et al. 2016; Marshall and Stepien 2019), whereas others did not (Shaw et al. 2016; Gillet et al. 2018). Biases due to the degree of match between primers and target sequences can significantly affect these relationships and/or inhibit species detections (Xiong et al. 2016; Alberdi et al. 2017; Kelly et al. 2017). Some general markers have used less variable gene regions such as mitochondrial (mt) 12S rRNA to facilitate better match between primers and target sequences, which often limit resolution to the genus level or higher (Miya et al. 2015; Valentini et al. 2016; Cilleros et al. 2019). Metabarcoding results also have been used to evaluate population genetic information, employing specifically-designed, targeted markers to amplify sequence regions containing intra-specific haplotypes (Sigsgaard et al. 2017; Parsons et al. 2018; Marshall and Stepien 2019; Stepien et al. 2019).

### Potential error sources in environmental metabarcoding assays

PCR inhibition is a challenge in some environmental samples, leading to amplification failure or false negatives (Civade et al. 2016; Fujii et al. 2019). Error from incorrect base calls and/or sequence-to-sample mis-assignments due to index-hopping (when the wrong index is incorporated into a HTS library) can result in false positives (Xiong et al. 2016; Deiner et al. 2017). Sequencing error and index-hopping particularly are problematic in discerning invasive or rare species, since a false positive can lead to wasted effort and funds, including unnecessary response by management agencies to verify presence

(Zaiko et al. 2018). More stringent bioinformatic filtering can remove some of this error but may lower detection capability, particularly for rare taxa. Better protocols are needed to alleviate primer bias and error in assay data, while correctly identifying as many taxa as possible (Zinger et al. 2019). Inaccurate base calls in HTS can artificially inflate population diversity estimated from intra-specific haplotypic diversity (Tsuji et al. 2018), which may pose difficulty in distinguishing signal (correct haplotypes) from noise (“false” haplotypes).

### Objectives

Our research objectives were to: (1) test the use of multiple metabarcoding assays and an associated bioinformatic pipeline, which combined results from primer sets to reduce possible sources of error and increase confidence, and (2) evaluate the efficiency and accuracy of this approach in field and laboratory experiments. For (2), we compared the results with those from traditional capture-based field sampling of fishes, morphological identifications, and population genetic Sanger sequencing of individuals.

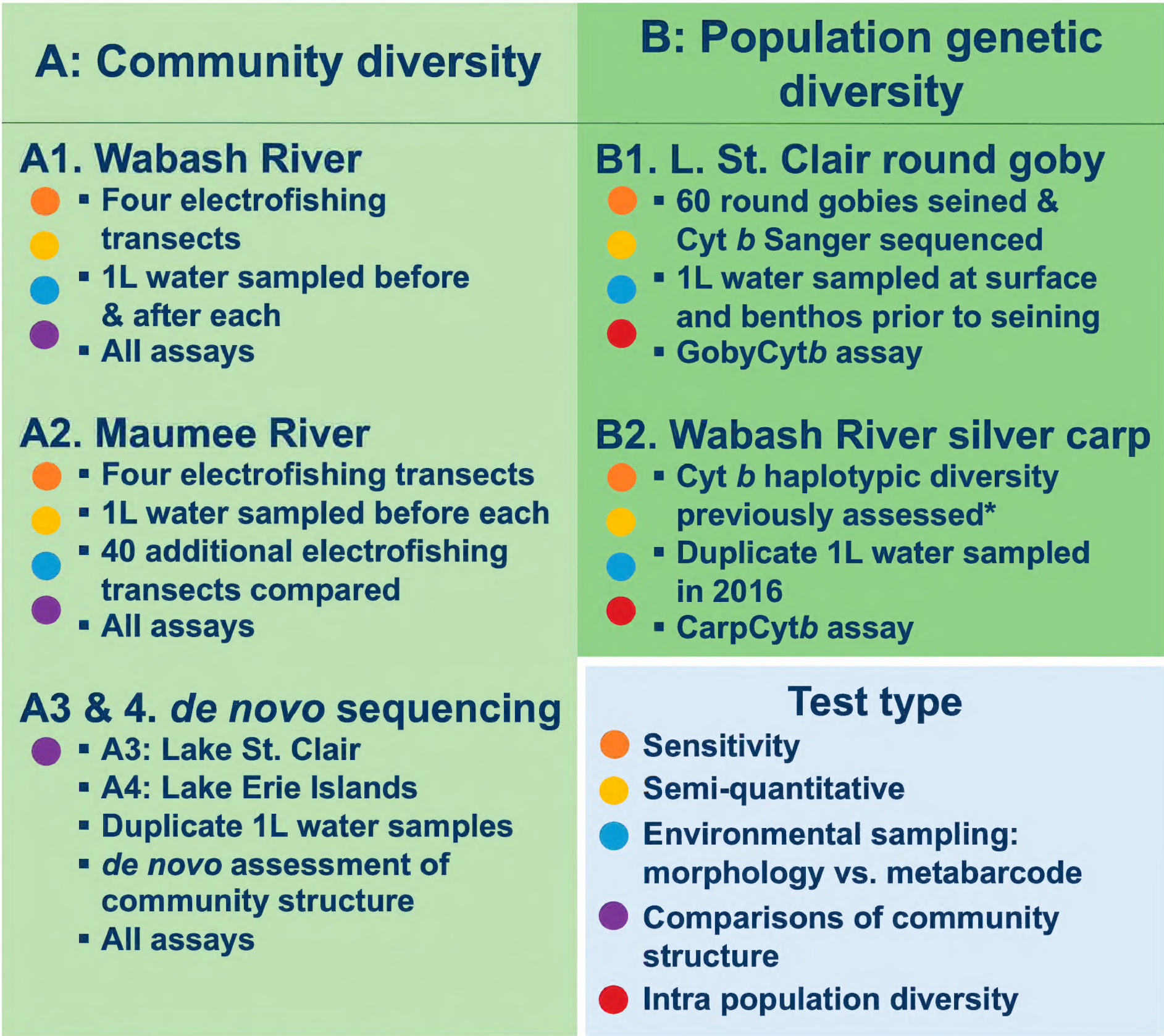
We tested the performance of our metabarcoding assays and bioinformatic pipeline with mock communities, laboratory aquarium experiments, and processing of an ichthyoplankton sample to assess sensitivity for assessing inter- and intra-specific diversity (Suppl. material 1). We applied this metabarcoding protocol to eDNA water samples from two large rivers (Figs 1, 2), one in the Mississippi River system (the Wabash River; Experiment A1) and the other in the Great Lakes’ watershed (the Maumee River; Experiment A2). These were taken concomitant with electrofishing surveys and with *de novo* sequencing of fish community eDNA from two Great Lakes’ sites, in Lake St. Clair and Lake Erie (Experiments A3–4; Figs 1, 2). In addition, we conducted further field experiments to assess the ability of metabarcoding assays to discern population (haplotypic) genetic diversity (Experiment Series B; Figs 1, 2).

## Methods

### Ethics statement and fish sampling

All fishes were collected by our lab under Ohio Department of Natural Resources (ODNR) permit #17-159, Michigan Department of Natural Resources permits, or by collaborators with their permits (see Acknowledgements). All native fishes except those used for the mock communities (Suppl. material 1) were released alive and in apparent good health immediately in the sampling area. Invasive fishes (which cannot be legally re-released) were anesthetized and sacrificed under the approved University of Toledo IACUC #205400, “Genetic studies for fishery management” (to CAS and laboratory members) using an overdose of 250mg/L tricaine methane sulfonate (MS222; Argent Chemical Laboratories,





**Figure 1.** Experimental design schematic, depicting Experiment Series A and B, brief methods summary for each experiment in the Series, the aspect of metabarcoding assays tested, and assays applied.  
\* silver carp haplotypic diversity was assessed by Stepien et al. (2019).

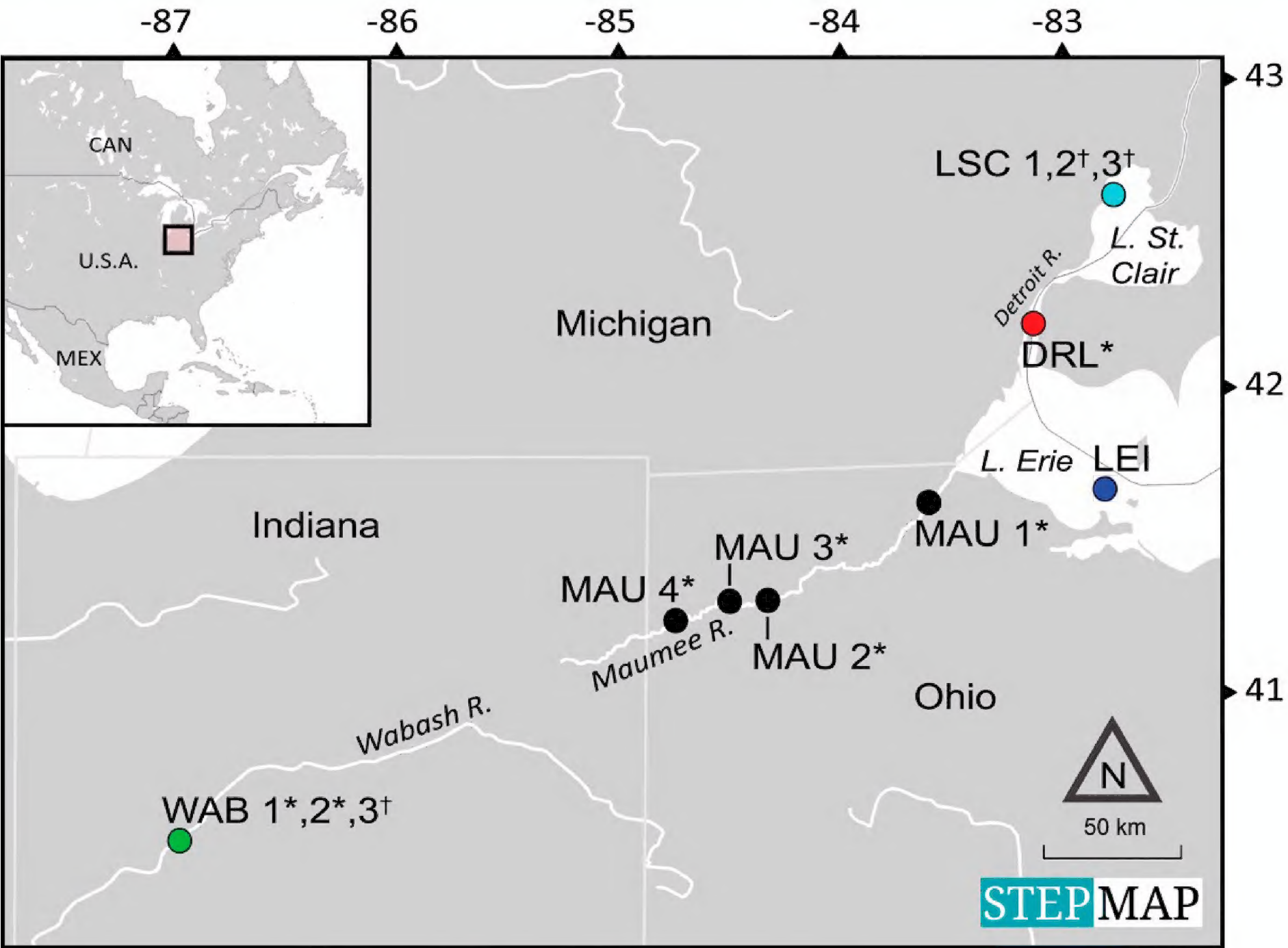
Redmond, WA). Taxonomy and nomenclature presented followed [www.Fishbase.org](http://www.Fishbase.org).

Metabarcoding assays employed

Three metabarcoding assays designed by our lab (Stepien et al. 2019, Snyder et al. 2020) were used, which targeted the mtDNA cytochrome (cyt) *b* gene to identify and differentiate fish species (native and introduced/invasive) from the Great Lakes and upper Wabash River regions (Table 1, Suppl. material 1: Fig. S1). The FishCyt*b* assay was a general assay designed to detect all fishes in these ecosystems (Snyder et al. 2020), which amplified a 154 nucleotide (NT) region of the cyt *b* gene, beginning at NT 855. The CarpCyt*b* assay was formulated for invasive carps (Stepien et al. 2019), amplifying 136 NTs beginning at NT 114, and the GobyCyt*b* assay (Snyder et al. 2020) was designed to distinguish invasive gobies,

amplifying 167 NTs beginning at NT 42. Another general fish assay for part of the mt 12S RNA gene (MiFish; Miya et al. 2015) also was used, for comparison (Table 1). The CarpCyt*b* and GobyCyt*b* assays also were designed to detect much of the known population genetic haplotypic diversity across the target taxon group’s respective invasive North American range.  
All primer sets included the Illumina sequencing adapters and four unique spacer inserts, designated e–h, at the 5’ end (Table 1; Klymus et al. 2017). Spacer inserts varied from 7–14 NTs to offset sequences and increase library diversity, thereby improving the quality of HTS data on the Illumina platform (Fadrosh et al. 2014; Wu et al. 2015). The assays and their associated bioinformatic pipeline were tested on mock communities (Experiment Series C1), in which samples containing genomic extractions of taxa having known sequences for the respective markers were mixed in a factorial design of varying





**Figure 2.** Map showing sample sites in the Wabash River (WAB), Maumee River (MAU), Detroit River (larval fish sample; DRL), Lake St. Clair (LSC), and Lake Erie Islands (LEI) (for Experiment Series A and B). At selected sites, morphological surveys (\*) or traditional population genetics sampling and data collection (†) were conducted and compared to eDNA metabarcoding assay results. Wabash River (WAB) and Lake St. Clair (LSC) locations were in too close proximity to be depicted separately (geographic coordinates are in Suppl. material 1: Table S1). Field locations were mapped using STEPMAP (stepmap.com, which holds no copyright on data or layers presented).

**Table 1.** Primers used for our metabarcoding assays. Table indicates primer element function, primer name, direction (Direction (Dir); F=forward, R=reverse), and sequences for each primer element. Length of region amplified (NTs; variable for 12S RNA MiFish, for which a mean is given) and annealing temperatures ( $T_A$ ) are provided for target-specific primers. Primer topology was 5'–Illumina sequencing adapter, spacer insert, target specific primer–3'. Spacer inserts were from Klymus et al. (2017). Assay publications: *CarpCytb* (Stepien et al. 2019), *GobyCytb* and *FishCytb* (Snyder et al. 2020), and 12S RNA MiFish (Miya et al. 2015).

Function	Name	Dir	Sequence 5'–3'	NTs	$T_A$
Target specific	FishCytb	F	GCCTACGCTATYCTHCGMTCHATYCC	154	50 ° C
		R	GGGTGTCNACNGGYATNCCNCCAATTCA		
	CarpCytb	F	KRTGAAAYTTYGGMTCYCTHCTAGG	136	54 ° C
		R	AARAAGAATGATGCYCCRTTRGC		
	GobyCytb	F	AACVCAYCCVCTVCTWAAAATYGC	167	50 ° C
		R	AGTCANCCRAARTTWACRTCWCGRC		
	MiFish	F	GTCGGTAAAACTCGTGCCAGC	~172	65 ° C
		R	CATAGTGGGGTATCTAATCCCAGTTTG		
Adapter	Illumina seq	F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG		
		R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG		
Spacer inserts	e	F	TCCTATG		
		R	CGTACTAGATGTACGA		
	f	F	ATGCTACAGT		
		R	TCCTAGCTGACGC		
	g	F	CGAGGCTACAACTC		
		R	GAGTAGCTGA		
	h	F	GATACGATCTCGCACTC		
		R	ATCGGCT		



dilutions (per Klymus et al. 2017; Marshall and Stepien 2019). We also conducted aquarium experiments (Experiment Series C2), which tested detection of both inter- and intra-specific variation, and evaluated a bulk ichthyoplankton sample from which species had been identified using morphological characters and microscopy (Experiment C3). (These are detailed in Suppl. material 1: Experiments C1–C3).

### Experiment Series A: Species compositions and community diversity

To compare among the assays and with results from traditional morphological identifications of capture-based samples, several eDNA water samples were collected concomitant with conventional surveys (Figs 1, 2, Suppl. material 1: Table S1). Samples were taken in bleach-sterilized 1 L bottles 10 cm below the water surface, stored on ice in brief transit from the field, and then frozen at -80 °C. These included duplicate 1 L eDNA water samples collected before and after two electrofishing transects conducted by us on September 2, 2016 in the Wabash River, IN (Experiment A1, sites WAB 1–2 in Fig. 2), where invasive silver carp *Hypophthalmichthys molitrix* was prevalent and bighead carp *H. nobilis* also was present.

The Ohio Environmental Protection Agency (OEPA) conducted 44 electrofishing surveys at 22 sites in the Maumee River, OH (a western Lake Erie, Great Lakes tributary) in June–September 2012, from which all fishes were identified, counted, and weighed (g) by them (OEPA 2014, 2015). Immediately prior to their sampling, they collected water samples for us 10 cm below the surface in sterile 1 L containers, of which four sites were analyzed here (Experiment A2, sites MAU 1–4 in Fig. 2). We also analyzed 1 L eDNA water samples collected in duplicate by us from Lake St. Clair at the Harley Ensign Memorial Boat Launch, Harrison Charter Township, MI on June 5, 2017 (Experiment A3, site LSC 1 in Fig. 2) and from Lake Erie at the Franz Theodore Stone Laboratory, Put-In-Bay, Gibraltar Island, OH on July 29, 2017 (Experiment A4, site LEI in Fig. 2). Experiments A3 and A4 involved *de novo* sequencing of eDNA water samples without accompanying morphological survey data, aimed to further test the ability of these metabarcoding assays to differentiate among fish communities from various habitats.

### Experiment Series B: Population genetic variation with metabarcoding assays

To evaluate population genetic compositions using metabarcoding assays of eDNA water samples, 1 L of water was collected 10 cm below the surface (site LSC 2 in Fig. 2) and 10 cm above the benthos (LSC 3), immediately prior to seining 60 round gobies on November 16, 2016 from the Lake St. Clair sampling site analyzed for B1 (Figs 1, 2, Suppl. material 1: Table S1), from which complete *cyt b* sequences were obtained

following Snyder and Stepien (2017). *Cyt b* sequence haplotypes and population genetic diversity of the silver carp previously were assessed from 37 individuals from a Wabash River site (WAB 3 in Fig. 2) by Stepien et al. (2019) and were further analyzed here using duplicate 1 L eDNA water samples collected 10 cm below the surface, as above, on September 2, 2016 (Experiment B2, Figs 1, 2). Traditional population genetic sequencing results were compared with those from eDNA water samples using the *CarpCytb* and *GobyCytb* assays.

### DNA capture, extraction, and library preparation

Water samples from turbid habitats often clog filters (Williams et al. 2017) and thus genetic material was captured using centrifugation (per Stepien et al. 2019; Snyder et al. 2020). The DNA was extracted from the pellets using Qiagen DNeasy kits (Hilden, Germany) (detailed in Suppl. material 1). A custom library preparation (prep) protocol (Stepien et al. 2019; Snyder et al. 2020) included extraction controls, centrifugation controls, no-template PCR controls (for initial amplifications and index reactions), and PCR clean-up controls. Positive DNA controls (Suppl. material 1) were used to quantify and correct for potential errors, which might include incorrect base calls, index-hops (see “*Bioinformatic pipeline*” below), and/or cross-contamination of samples. The positive controls contained equal mass (ng) of DNA from 10 marine fish species that could not survive in freshwater, which had been Sanger sequenced by us (Suppl. material 1: Table S2). Unexpected sequences of positive control taxa in the metabarcoding results (those not present in the Sanger sequencing results) were used to calculate an error rate (see “*Bioinformatic pipeline*” and Suppl. material 1). This library prep protocol also utilized primers with custom spacer inserts, with the latter serving as first step PCR indices to facilitate detection and removal of cross-contamination and index-hopping. The latter can occur when the wrong index is incorporated into a HTS library, leading to sequence-to-sample mis-assignment (Xiong et al. 2016; MacConaill et al. 2018).

### Bioinformatic pipeline

We used our previously published custom bioinformatic pipeline (Stepien et al. 2019; Snyder et al. 2020). Primers were trimmed from raw reads using a custom Python v3.7.1 script, which allowed for sequencing errors in 30% of the primer NTs (a standard approach for metabarcoding assays; see Deiner et al. (2017)). Since PCR errors in positive controls tended to occur at the NT positions immediately following the forward or reverse primers, those first and last NTs also were trimmed (Suppl. material 1). The trimming script additionally removed any reads having the wrong spacer insert, which may result from index-hopping or cross-contamination, and eliminated non-informative sequences < 100 NTs from primer dimers (Khodakov et al. 2016).



Trimmed reads were merged in DADA2 (Callahan et al. 2016), which corrected sequence errors using a de-noising algorithm and removed chimeras. We used DADA2 default parameters, with “MaxE” set to “(3, 5)”. Inputs were truncated at median Q score < 30, for the first 10 samples/assay/run, assessed using DADA2’s plotQualityProfile function. De-noised sequences having 100% similarity in DADA2 are termed amplicon sequence variants (ASVs).

Unique ASVs were subjected to the Basic Local Alignment Search Tool (BLAST; <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) from the command line against custom databases, to obtain the top 500 results per ASV. The custom database consisted of all cyt *b* or 12S RNA (MiFish) sequences on GenBank for fishes from the Great Lakes, plus predicted future invasive species (NOAA 2019). These databases were curated using GenBank searches for “cytochrome *b*” or “12S RNA” and “Actinopterygii”, all sequences were downloaded in FASTA format, and then those species that were not present (or predicted future invaders) in the systems were removed (Hubbs and Lagler 2007). ASVs of samples taken from the Wabash River were subjected to BLAST searches against databases containing all available Actinopterygii fish cyt *b* or 12S RNA sequences.

Our cyt *b* reference database was robust, containing sequences for > 95% of Great Lakes fishes and 100% of the present and predicted invasive species (see Snyder et al. 2020). Cyt *b* sequences for two native catostomid fishes, three extant *Coregonus* spp. and a believed-to-be extinct species in that genus, two cyprinids, and troutperch were absent from GenBank at the time of our study. In contrast, the Great Lakes 12S RNA reference sequence database on GenBank was missing 53 (23%) of known Great Lakes’ fish species. Sequences for 12S RNA were absent for many native taxa, notably 80% of *Coregonus* spp., 47% of catostomids, 22% of cyprinids, 21% of percids, and several others. Also absent were several current (i.e., round goby and tubenose goby) and predicted possible Great Lakes’ invasive species (i.e., steelcolor shiner *Cyprinella whipplei*, stellate tadpole goby *Benthophilus stellatus*, Black Sea monkey goby *Neogobius fluviatilis*, and Caspian Sea monkey goby *Neogobius pallasii*).

BLAST results (i.e., “hits”) with < 90% query cover or identity were removed. All unique species hits per ASV per assay that passed this filter and had the lowest expectation (e) value (best match) were combined in a list of potential taxa. A lowest common ancestor approach was used for taxonomic assignments for sequences that did not match a single species (i.e., if all hits with the lowest e value were the same genus, then taxonomic assignment was to that level).

For metabarcoding assay results, species incidences were scored as valid when they were greater than the calculated error cutoff from the positive controls (see “DNA capture, extraction, and library preparation” above and Suppl. material 1) in a single assay or occurred in multiple assays (hereafter termed the multi-assay approach). This approach aimed to reduce false positives and simul-

taneously compensate for potential primer set biases. We compared those results to the use of 0.1% as the cutoff (the known rate of index-hopping on MiSeq; MacConaill et al. 2018), and also evaluated all positive hits (without frequency-based filtering, but passing the % identity and query cover filter). Metabarcoding sequence data for replicates from the same sampling site were combined, by applying the bioinformatic filter for each separately and then by combining ASVs and read counts afterwards (independently for each primer set). Species detections per assay and for the combined assays (multi-assay approach) were compared to results from the accompanying morphological capture-based sampling. Scripts for this bioinformatic pipeline and custom BLAST databases are deposited in the Dryad database (<https://doi.org/10.5061/dryad.7m0cfxprx>). FASTQ files for all samples sequenced are in the NCBI Sequence Read Archive (BioProject Accession: PRJNA625378).

### Data analyses

Results from our metabarcoding assays were compared with their concomitant morphological identifications from traditional capture-based surveys (i.e., electrofishing transects A1 and A2, or Sanger sequencing of sampled fish in B1 and B2). Species appearing unique to the metabarcoding assays and/or capture-based surveys were evaluated from individual samples and sampling regions. Some morphological or metabarcoding identifications were restricted to the genus or family levels (see). False negatives were determined using a relaxed detection criterion, which recorded a species as being present when distinguished at the genus level in either the morphological or metabarcoding results. Species richness values were compared between the morphological and the metabarcoding approaches (and among the individual and multi-metabarcoding assays) using *t*-tests in R (R Core Team 2018), with significance adjusted using sequential Bonferroni corrections (hereafter, SBC, Rice 1989). Biomass proportions of species from morphological identifications were compared statistically to the proportions of sequence reads for single metabarcoding assays using linear models and Spearman rank correlation coefficients in R. Analysis of Covariance (ANCOVA) compared the slopes of the relationships among the different assays to evaluate concordance (Zar 2010).

Habitat differentiation was assessed with Non-metric Multi-Dimensional Scaling (NMDS) using Bray-Curtis dissimilarity in VEGAN (Oksanen et al. 2019) for morphological identifications and individual and multi-metabarcoding assays, based on presence/absence (binary) analysis. Differences in species compositions between the Wabash (A1) and Maumee rivers (A2) were statistically tested using ANOVA, with the ADONIS2 function in VEGAN. The assumption that groups of points did not significantly differ in their distance to the centroid was tested using BETADISPER. Habitat comparisons further were explored using FASTCLUSTER in R. Metabarcoding



assay and capture-based survey dendrograms were constructed using binary distances (based on presence/absence of taxa) and Ward’s D2 agglomeration method (Müllner 2013). PVCLUST calculated approximate unbiased (AU) percent bootstrap support for each node in the dendrogram (10,000 replications), using the same distance and agglomeration methods (Suzuki and Shimodaira 2015).

Numbers and proportions of population haplotypes were calculated from HTS reads and traditional Sanger sequencing of individuals.  $F_{ST}$  and exact tests of population differentiation in Arlequin (Excoffier and Lischer 2010) compared the proportions of haplotypes found with traditional and metabarcoding methods.

Results

High-throughput sequencing metrics

No PCR amplifications occurred in the 0 hr control samples from the goby aquarium experiments (C3) or in any of our negative extractions, centrifugations, no-template PCR, indexing, or clean-up controls (see Suppl. material 1). A total of 27,961,011 sequence reads were obtained among all libraries (mean per sample per assay  $\pm$  SE=229,189  $\pm$  18,645; Suppl. material 1: Table S3). A mean of 204,320  $\pm$  17,738 reads per sample per assay was successfully trimmed. DADA2 merged an average 0.80  $\pm$  0.01 of trimmed reads, with a mean of 75.4  $\pm$  11.4

ASVs per sample per assay. Of those, a mean of 23  $\pm$  1.7 had BLAST hits to our fish databases that passed the identity and query cover filter of > 90% (mean query cover = 99.83  $\pm$  0.01%, mean identity = 99.14  $\pm$  0.02%). Single species identifications included 2,342 ASVs (89% of total 2,631 hits passing the filter) and 59, 103, 43, and 75 of the hits to the genus level for the FishCytb, CarpCytb, GobyCytb, and 12S RNA MiFish assays, respectively. These included 14 genera, primarily: *Carassius* (13% of the overall genus level hits), *Carpiodes* (19%), and *Ictiobus* (39%), for which either morphological identifications and/or another one of our metabarcoding assays resolved a congener to species. Nine 12S RNA hits were resolvable only to the family level, of which six were Cyprinidae, and three were Catostomidae; all were discarded. Not all eDNA samples led to successful libraries for every assay, presumably due to primer-specific inhibition (dashes in Table 2). For all positive controls, the most abundant unexpected sequence was closely related to an expected sequence. Error frequencies calculated from the positive controls ranged from 0.18–0.42% (mean = 0.27  $\pm$  0.02%).

Experiment Series A: Metabarcoding assays versus morphological identifications

Laboratory experiments and analyses of the larval fish sample showed that our assays and accompanying bioinformatic pipeline were highly sensitive, with few false negatives and high detection probability (Suppl. mate-

**Table 2.** Sample diversity based on morphological identifications versus metabarcoding results (from Experiment Series A and B). (A) Species richness from morphological surveys and metabarcoding assays. (B) Number of species (richness) discerned with morphology and species uniquely found with metabarcoding results (morphological “false negatives”). Proportion of false negatives in metabarcoding results (in parentheses). Regional samples were combined, respectively, for the Maumee (1–4) and Wabash (1–2) rivers. For the Maumee River, “all” indicates results for all species in summer 2012 electrofishing surveys regardless of whether their concomitant eDNA data were processed.

A		Richness				
Location	Morphology	Multi-assay	FishCytb	CarpCytb	GobyCytb	12S RNA MiFish
Maumee River 1	13	38	19	14	18	18
Maumee River 2	22	25	15	12	9	7
Maumee River 3	23	26	18	10	6	16
Maumee River 4	23	28	17	6	15	19
Maumee River 1–4	33	42	26	20	24	31
Lake St. Clair 1	–	16	6	7	8	8
Lake St. Clair 2	–	16	–	10	12	12
Lake St. Clair 3	–	16	9	11	9	–
<i>Lake St. Clair all</i>	–	23	6	16	16	8
Lake Erie Islands	–	14	7	8	–	9
Wabash River 1	13	30	8	14	–	21
Wabash River 2	12	29	14	10	5	16
Wabash River 3	–	27	17	11	9	11
Wabash River 1–2	18	37	22	20	10	30
B		Unique to metabarcoding assays (false negatives)				
Location	Richness Morphology	Multi-assay	FishCytb	CarpCytb	GobyCytb	MiFish
Maumee River 1	13	20 (0.00)	1 (0.23)	1 (0.38)	3 (0.23)	7 (0.31)
Maumee River 2	22	6 (0.41)	3 (0.45)	3 (0.73)	1 (0.73)	1 (0.77)
Maumee River 3	23	4 (0.26)	0 (0.57)	1 (0.70)	0 (0.78)	3 (0.52)
Maumee River 4	23	10 (0.30)	1 (0.61)	1 (0.83)	4 (0.65)	6 (0.57)
Maumee River 1–4	33	18 (0.12)	2 (0.36)	4 (0.67)	6 (0.58)	9 (0.39)
<i>Maumee River all</i>	59	9 (0.27)	0 (0.56)	2 (0.76)	3 (0.64)	6 (0.54)
Wabash River 1	13	16 (0.23)	0 (0.54)	5 (0.54)	–	13 (0.46)
Wabash River 2	12	14 (0.08)	7 (0.67)	1 (0.50)	2 (0.83)	7 (0.17)
Wabash River 1–2	18	21 (0.00)	9 (0.33)	5 (0.39)	2 (0.83)	14 (0.22)



rial 1). Several false negatives, i.e., taxa discerned by morphological identifications but not with single metabarcoding assays using the calculated error cutoff, were positive when 0.1% was used for filtering ( $N = 13$ ) or when all ASVs were accepted ( $N = 48$ ). Our multi-assay approach detected more species than did the single assays (see “Community comparisons” below). Using the multi-assay approach, just one false negative (with the calculated error cutoff) was positive when all ASVs were accepted. When ASVs above 0.1% were accepted, several index-hops were apparent, including for the Black Sea sprat *Clupeonella cultriventris*, a possible future invader of the Great Lakes that has not been documented in North America (NOAA 2019), and for silver carp outside of its known established invasive range in the Mississippi River basin (Kolar et al. 2005). Those likely mis-assigned (index-hops) from mock communities on the same run or resulted from cross-contamination, since both had been previously Sanger sequenced in our lab. Hits for cod *Gadus* spp. and rockfish *Sebastes* spp., marine taxa that were used as positive controls here, also appeared in some metabarcoding results from eDNA samples, when 0.1% was used as the frequency-based filtering cutoff. Thus, we used the calculated error cutoff due to the possibility of false positives under less stringent filtering conditions.

Morphological identifications from capture-based surveys did not completely overlap the metabarcoding assay results. For the total of all samples, our multi-assay metabarcoding approach detected more taxa than did morphological determinations (Table 2A). Results revealed that 51 taxa (74%) were in common between both approaches in the Wabash (A1) and the Maumee (A2) rivers (Suppl. material 1: Table S4). Since hybrid species identified morphologically (i.e., hybrid striped bass *Morone chrysops* x *saxatilis*) possess a single mtDNA genome, our metabarcoding assays discerned the maternal species alone. Metabarcoding assays identified some samples to species that were morphologically unresolved. Several metabarcoding false negatives were restricted to the genus level, either due to lack of taxonomic resolution or incomplete BLAST databases. With those corrections, 93% of the species detected in capture-based surveys (taken concomitant with eDNA water sampling) also were identified by the metabarcoding assays. There was no relationship between degree of primer sequence match and false negatives (Suppl. material 1: Fig. S1). A positive (but weak) correlation was observed between biomass of species and sequence reads (Suppl. material 1: Table S5, Fig. S2). This also was not significantly related to the degree of primer sequence match.

Experiment A1: Morphology discerned 18 fish species belonging to five families, from electrofishing surveys in the Wabash River. Our metabarcoding assays identified all of those (100%) to species, along with an additional 20 species (Fig. 3, Table 2B, Suppl. material 1: Table S4).

Experiment A2: 33 species belonging to 11 families were detected with electrofishing surveys conducted concomitant with eDNA water sampling from four sites in

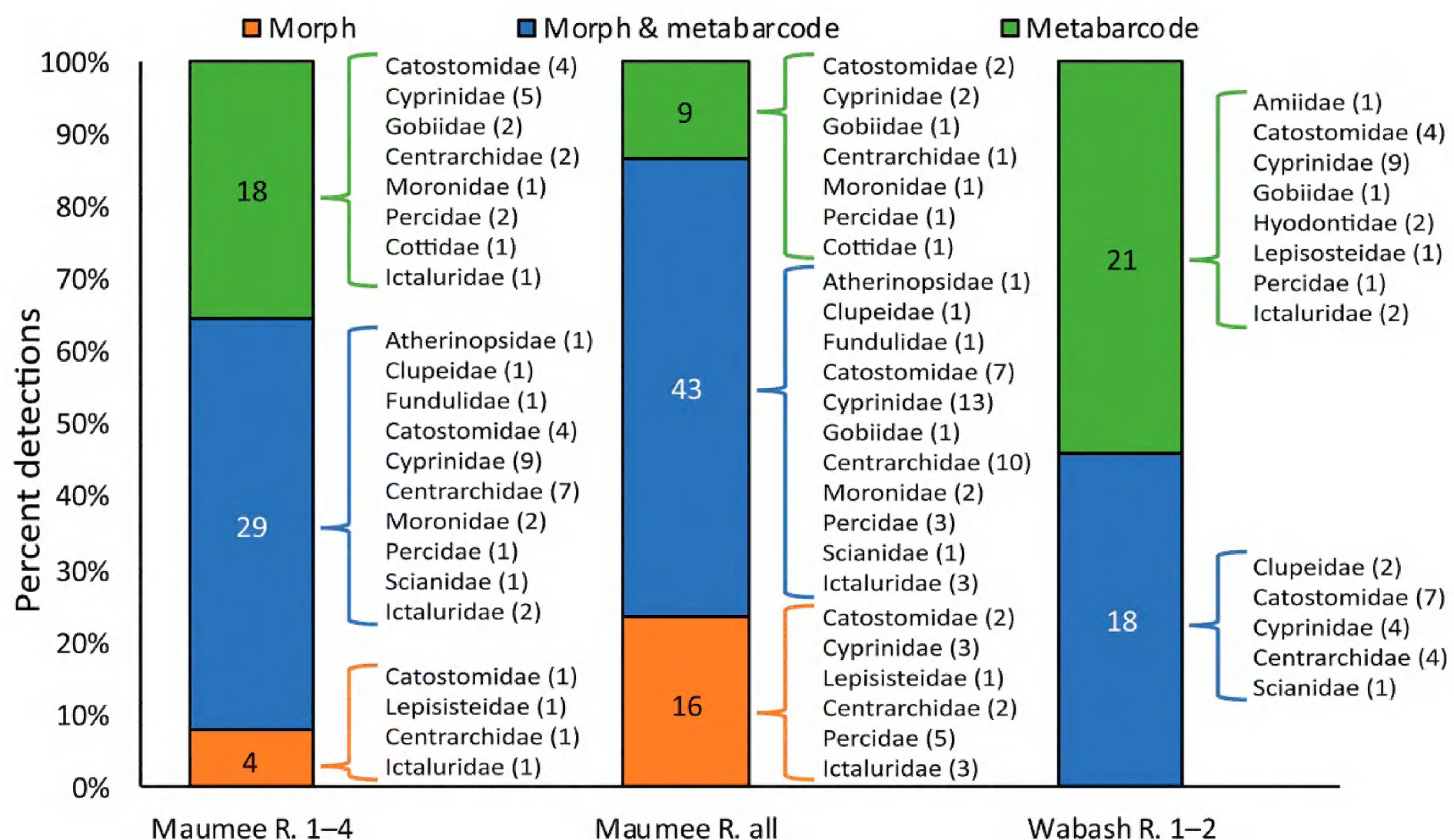
the Maumee River. Our metabarcoding analyses detected 29 (88%) of those, along with an additional 19 species. A total of 59 species belonging to 12 families were collected among all 44 morphological surveys from 22 Maumee River sites across four months of intensive sampling by the OEPA during summer 2012. Our metabarcoding assays discerned 43 (73%) of those species and an additional nine species from just single 1 L water samples at only four of the sites (corresponding to 9% of the OEPA’s surveys, and 18% of their total number of sampling sites).

Only four of the fish species from the Maumee River electrofishing surveys conducted concomitant with our four single eDNA water samples were not detected with metabarcoding assays – northern hogsucker *Hypentelium nigricans*, longnose gar *Lepisosteus osseus* (the sole false negative in our high diversity aquarium experiments; Suppl. material 1: Experiment C2), stonecat *Noturus flavus*, and white crappie *Pomoxis annularis* (Fig. 3, Table 2B, Suppl. material 1: Table S4). Thirteen additional species uncovered in the Maumee River across the OEPA’s entire four month-long morphological electrofishing survey did not occur in metabarcoding results from the four water samples (Table 2B), including rock bass *Ambloplites rupestris*, golden redhorse *Moxostoma erythrurum*, five darter species, four small cyprinids, and two madtom (*Noturus*) species.

As expected, species results from the single assays did not completely overlap. Of the 347 individual species detections across all samples, 111 (32%) occurred in single assays. Twenty-one (6%) of the detections were scored as positive according to the multi-assay criteria, meaning that their hits fell below the cutoff values for multiple assays in the same sample. Mean proportions of false negatives from single assays in samples taken concomitant with electrofishing surveys were  $0.48 \pm 0.04$ . When all samples from a single region were combined, this value fell to  $0.34 \pm 0.04$ . The multi-assay approach had significantly fewer false negatives after SBC than did the single assays ( $p < 0.004$  for all). Mean proportions of false negatives using the multi-assay approach were  $0.17 \pm 0.05$  for the individual sampling sites, and  $0.09 \pm 0.03$  when all samples from each region were combined. Six of the common false negatives from the 12S RNA assay were due to species lacking reference 12S RNA sequences in GenBank (i.e., quillback *Carpiodes cyprinus*, highfin carpsucker *Ca. velifer*, shorthead redhorse *Moxostoma macrolepidotum*, ghost shiner *Notropis buechanani*, and white crappie).

Complete (100%) detection from metabarcoding assays occurred for Experiment A2 in the MAU 1 sample, which had the lowest morphological species richness (Table 2B). Comparing the four eDNA water sample metabarcoding results to the entire scope of the electrofishing surveys conducted by the OEPA in the Maumee River throughout summer 2012, each single assay detected between 26% (Carp-Cytb) to 48% of the overall species (12S RNA MiFish). The multi-assay approach discerned 73% of the species overall in this watershed level community, based on just four 1 L water samples (one each from four different sites).





**Figure 3.** Families (and numbers of fish species, in parentheses) detected with morphology (Morph), metabarcoding assays, or using both methods (for Experiment Series A). Samples taken concomitant with electrofishing surveys were combined (Maumee River 1–4, Wabash River 1–2). Maumee River all: comparison of four eDNA water samples to 44 electrofishing transects from 22 sites in the Maumee River, June–September 2012.

A mean of  $4.6 \pm 1.0$  taxa in the single assays or  $14.5 \pm 5.3$  using the multi-assays were undetected in the concomitant morphological samples. Two unlikely false positives occurred with the 12S RNA MiFish assay. There were several apparent matches to the non-native blacktip jumprock *Moxostoma cervinum* in the Wabash River, likely due to most *Moxostoma* being absent from the 12S RNA database – six of the seven species known from the watershed (Simon 2006) lacked reference sequences. Instead, other *Moxostoma* spp. were identified by the cyt *b* assays and/or morphology in every sample for which the 12S RNA assay displayed a hit for blacktip jumprock. The false hits from the 12S RNA assay were discarded from the final dataset. The other apparent false positive from the 12S RNA assay was marine white croaker *Genyonemus lineatus* (native to the Eastern Pacific and not known to be able to survive in fresh water), which was not present in positive controls. Just one species in that family is known to occupy the sampled regions – the freshwater drum *Aplodinotus grunniens*. The drum had a single 12S RNA reference sequence in the BLAST database, which multiple 12S RNA ASVs from other samples matched. Without a reasonable explanation, hits for that species also were removed from the final dataset.

The metabarcoding assays found every invasive species collected in the morphological capture-based surveys, including: silver carp, common carp *Cyprinus carpio*, flathead catfish *Pylodictis olivaris*, round goby,

and white perch (Suppl. material 1: Table S4). Ghost shiner eDNA was not detected from two Maumee River sites (A2) where it was physically collected, but was found in metabarcoding assay results from another sample in the region. Both of those false negatives occurred at  $< 0.1\%$  of the total fish biomass. Our assays identified more samples that contained invasive species. For example, just one electrofishing transect in the Wabash River (A1) caught silver carp, yet every metabarcode sample detected that species in at least one assay. Our assays detected invasive grass carp in the Maumee (A2) and Wabash rivers (A1), where it is known to occur but was not caught. Tubenose goby was not captured in our Maumee River samples (A2), but was present in the eDNA metabarcoding results (and is known to occur there).

### Experiment Series A: Community comparisons

Some species were detected in just one of the geographic regions we sampled. In the Wabash River (Experiment A1), these were: blue sucker *Cycleptus elongatus* and invasive silver carp *Hypophthalmichthys molitrix*, identified both with morphology and metabarcoding assays, and gravel chub *Erimystax x-punctatus* and mooneye *Hiodon tergisus* by eDNA metabarcoding assays alone. Species detected in the Maumee River samples alone were: pumpkinseed *Lepomis gibbosus*, orangespotted sunfish *Lepomis humilis*, invasive ghost shiner, spotted sucker



*Minytrema melanops*, and common logperch *Percina caprodes* with both morphology and metabarcoding assays, and black crappie *Pomoxis nigromaculatus*, spoonhead sculpin *Cottus ricei*, orangethroat darter *Etheostoma spectabile*, and invasive tubenose goby solely with metabarcoding assays (Experiment A2). We surveyed Lakes St. Clair (Experiment A3) and Erie (Experiment A4) with eDNA metabarcoding assays alone. Black bullhead *Ameiurus melas* solely was in the Lake Erie Islands sample, and invasive chum salmon *Oncorhynchus keta* in Lake St. Clair (where it has been introduced for sport fishing).

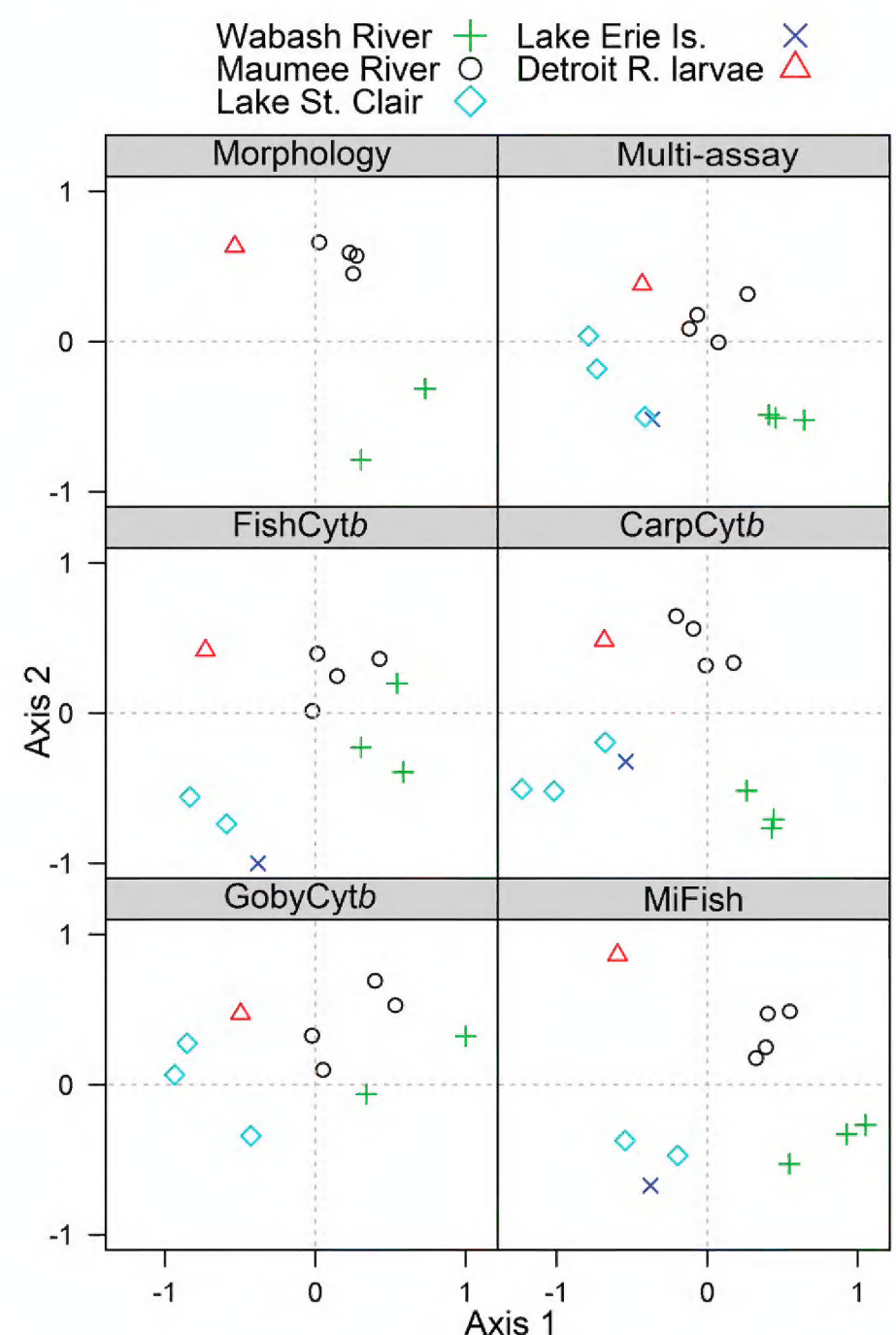
Species richness values obtained from single samples, as well as for regional analyses, were higher using the multi-assay approach than with any single metabarcoding assay or morphology (Table 2A). Richness values from morphology and metabarcoding assays were not significantly correlated. Richness values for the Wabash (A1) and Maumee (A2) rivers were statistically significantly greater with the multi-assay approach than for the other methods, including single assays and morphology ( $p < 0.004$  for all). Notably, numbers of taxa detected with the multi-assay approach that were undetected by morphological capture-based sampling, were greater than false negatives in all but two samples (MAU2–3; Table 2A). Numbers of replicates and/or samples collected per region were significant predictors of species richness for single metabarcoding ( $R^2 = 0.73$ ,  $p < 0.001$ ) and multi-assay results ( $R^2 = 0.79$ ,  $p < 0.001$ ).

NMDS plots discerned more discrete groupings of regional samples with multi-assays than with single assays (Fig. 4). Significant differences in distances to the centroid were not found for the Wabash (A1) and Maumee (A2) rivers with any method (morphology, individual assay, or multi-assays). Multi-assays ( $df = 1$ ,  $F = 6.32$ ,  $p = 0.030$ ) as well as the separate FishCytb ( $df = 1$ ,  $F = 3.00$ ,  $p = 0.031$ ), CarpCytb ( $df = 1$ ,  $F = 4.73$ ,  $p = 0.030$ ), and 12S RNA MiFish ( $df = 1$ ,  $F = 4.18$ ,  $p = 0.028$ ) assays resolved significance differences between the Wabash and Maumee river fish communities (A1 and A3), whereas the GobyCytb assay and morphological surveys did not.

Some samples did not cluster by geographic region in the dendrograms when single assay results were used together with morphological identifications (Fig. 5A). When the multi-assay approach and morphological data were used, all samples clustered according to region, showing improved resolution and site-specific discrimination (Fig. 5B). Lentic and lotic sites were well-differentiated. Fish communities from the two lake sites (Lakes Erie and St. Clair) were more similar to each other than either was to the river samples, in the multi-assays as well as in most single assays.

### Experiment Series B: Intra-specific population diversity from metabarcoding assays

Aquarium experiments using round gobies possessing haplotypes RG 1, 8, and 57 showed no false negatives,

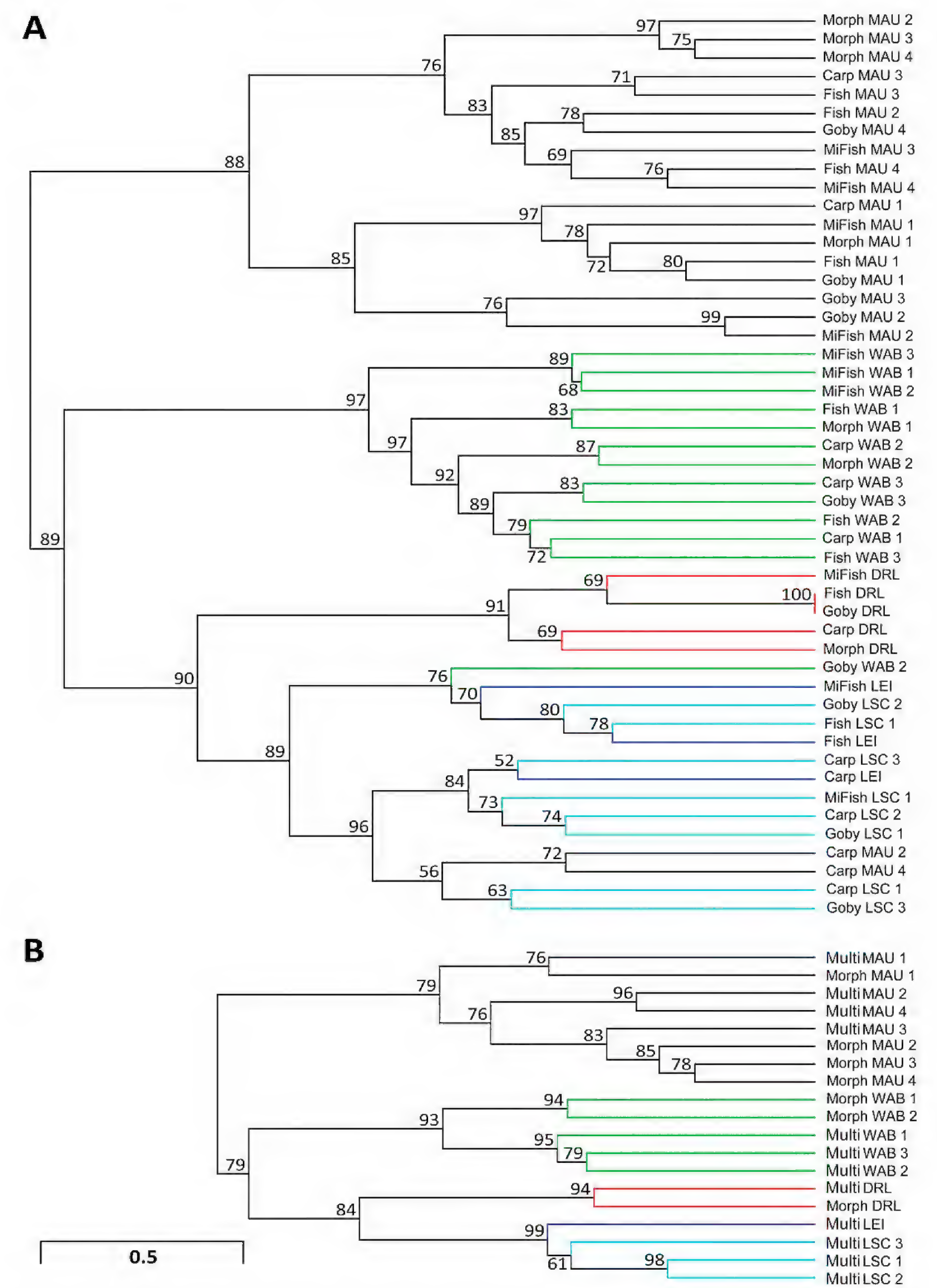


**Figure 4.** Non-metric multi-dimensional scaling plot based on binary Bray-Curtis dissimilarity for presence/absence of species detected by metabarcoding assays and morphological capture-based methods (when both were conducted concomitantly; Experiment Series A).

but some false positives fell above the error cutoff (Suppl. material 1: Experiment C3). Traditional Sanger sequencing of tissue samples identified three round goby haplotypes in Lake St. Clair (Experiment B1): “RG 1” (78% of individuals), “8” (12%), and “57” (10%) (Fig. 6). All three haplotypes were found in the benthic eDNA water sample processed with the GobyCytb assay. The two rare haplotypes were absent from surface water eDNA (even upon accepting all ASVs of any frequency or searching raw unmerged sequences). Both surface and benthic water samples contained multiple haplotypes that were above the calculated error cutoff, which were undetected by the Sanger sequencing analysis and were not in GenBank.

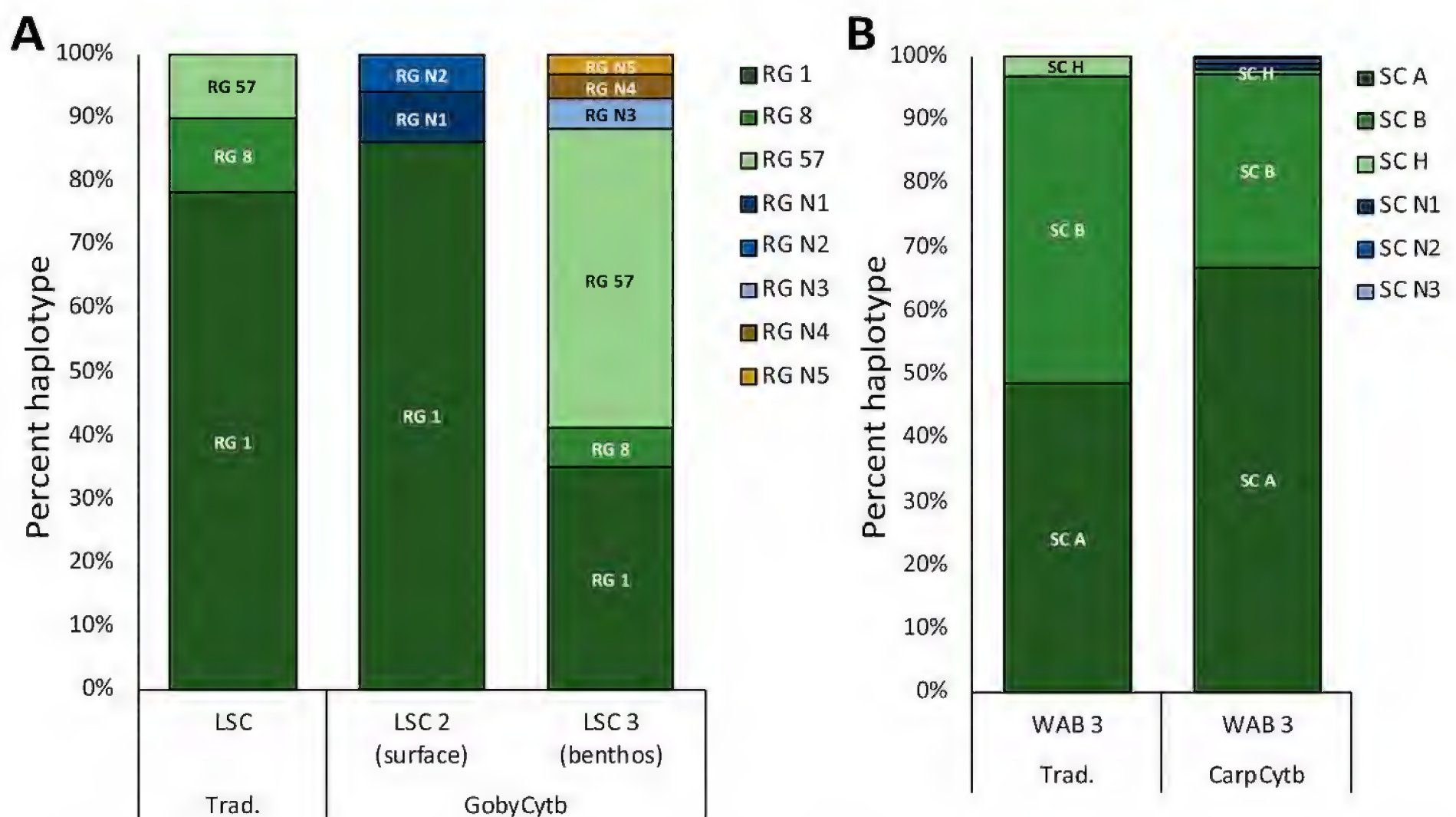
Sanger sequencing discerned three silver carp haplotypes that were physically sampled in the Wabash River (Experiment B2: designated as “SC A”, “B”, and “H”), constituting 49%, 48%, and 3% of that population sampled at a separate time (Stepien et al. 2019). The CarpCytb assay differentiated among all three of these haplotypes in the eDNA water samples, whose read proportions were 67%, 30%, and <0.5%, respectively, in a different year (Fig. 6). Three additional previously undiscovered haplo-





**Figure 5.** Dendrogram of relationships among metabarcoding and morphological samples, using binary distances and Ward's D2 agglomeration method (Experiment Series A). **(A)** Results from individual metabarcoding assays and morphological data. Fish = Fish-Cytb, Carp = CarpCytb, Goby = GobyCytb, MiFish = 12S RNA, Morph = morphological sampling. **(B)** Results from the multi-assay approach (Multi) and morphological (Morph) data. See Fig. 2 for color key and site abbreviations.





**Figure 6.** Population genetic haplotypic diversity assessed with metabarcoding assays versus traditional DNA sequencing (Experiment Series B). Round goby (RG) in Lake St. Clair (LSC2: surface, LSC3: benthos) and silver carp (SC) haplotypes in the Wabash River (WAB) assessed with traditional population genetic sequencing (Trad) and the GobyCytb and CarpCytb metabarcoding assays. New cytochrome *b* haplotypes (N) not described from either species to date, and having sequence frequencies <1% are unlabeled, for visual clarity.

types were detected – all above the calculated error cutoff and at greater frequency than the rare haplotype “H”. Comparisons of our metabarcoding assays with traditional population genetics based on Sanger sequencing of mtDNA haplotypes showed significant frequency differences after SBC using  $F_{ST}$  (mean  $F_{ST} = 0.179$ ,  $p < 0.0002$  for all) and exact tests ( $p < 0.0001$  for all).

## Discussion

Our multi-assay metabarcoding approach and accompanying bioinformatic pipeline demonstrated high detection probability that was better or similar to traditional morphological sampling, with low false negatives and additional species discerned despite considerably less sampling effort. The custom pipeline improved overall sequence run quality and removed apparent index-hops and/or cross-contamination by using spacer inserts, which served as indices for the initial amplifications. Sequencing error was calculated using positive controls of marine species that could not live in this freshwater environment. ASVs whose proportions were below the error cutoff were removed unless they occurred in multiple markers. Proportions of sequence reads showed weak but positive correlations to species biomass. Metabarcoding results for the round goby and the silver carp assays identified all of the haplotypic variation found with tradi-

tional population genetics Sanger sequencing. Additional “new” haplotypes found with metabarcoding assays may have resulted from sequencing error not removed by our pipeline, meriting further testing.

### Experiment Series A and B: Reducing error from metabarcoding assays

We evaluated various frequency-based filters to eliminate index-hops and/or cross-contamination from positive controls or other samples (since the likelihood that sequencing error would result in a BLAST hit to a different species was low). Given the large number of samples that can be pooled on a HTS run, such sources of error could result in false positives (Xiong et al. 2016; MacConaill et al. 2018). In our investigation, sequencing run quality was improved, and risks of cross-contamination or index-hops were reduced using the custom spacer insert library preparation protocol, which served to index the first step PCR. Our custom trimming script removed the majority of these errors. Despite the fact that the most common error in every positive control was closely related to an expected sequence (implicating sequencing error as the origin of these unexpected ASVs), our calculated error cutoff was the sole method that eliminated all apparent index-hops. Several studies have used positive controls to apply frequency-based bioinformatic filtering (Hänfling et al. 2016; Port et al. 2016; see Deiner et al. 2017). That approach likely



led to false negatives in our single assays. However, when the two targeted and the two general primer sets were combined in our multi-assay approach, false negatives and false positives were greatly reduced.

### Experiment Series A: Relative abundances of species from metabarcoding assays

Our results showed that the *cyt b* assays revealed strong positive relationships between input concentrations of DNA and the proportions of sequence reads in mock communities (Suppl. material 1: Experiment C1). eDNA water samples showed weaker, but most often positive relationships, depending on the assay used. These relationships likely were affected by environmental conditions, such as eDNA transport and settling rates in water (Deiner and Altermatt 2014, Pont et al. 2018), which vary under physical and biological conditions (Barnes et al. 2014; Jo et al. 2019) and whether the eDNA was intra- or extra-cellular/organellar (Turner et al. 2014). Variable abundances of mitochondria in different types of cells shed at different rates by different species also affect these relationships (Robin and Wong 1988; Klymus et al. 2017; Jo et al. 2019). Some authors have theorized that corrections for specific taxa, their sizes, life stages, and environmental conditions could be applied (van Bochove et al. 2020). Subsampling during collection and library preparation, as well as primer biases, also could have affected our results (Deiner et al. 2017).

Hänfling et al. (2016) discerned that metabarcoding assay read abundances were positively correlated with findings of long-term concomitant capture-based surveys of fishes in three United Kingdom lakes. Van Bleijswijk et al. (2020) found that fish biomass caught in fyke net surveys in a tidal inlet between the North and Wadden Seas showed weak positive correlation with metabarcoding sequence read abundances for just the eight most abundant species. Positive correlations with sequence reads at higher taxonomic levels have appeared more common in most studies (Thomsen et al. 2016; Gillet et al. 2018), although since different taxa in families often are not ecological equivalents (e.g., benthic invertivore darters and several piscivorous species in Percidae), those results are less useful than comparisons at the species level. Results here indicated that the selected assay influenced these relationships, as has been corroborated by other studies (Gillet et al. 2018).

### Experiment Series A: Resolution from metabarcoding assays versus conventional sampling

Elucidating overlap in community diversity between metabarcoding assays and traditional morphological capture-based survey methods was an important goal of our Experiment Series A. Other investigations have shown wide variation in species overlap between these approaches, ranging from 25% (Gillet et al. 2018; Cilleros et al. 2019) to > 90% (e.g., Hänfling et al. 2016, Port et al. 2016; Stoeckle

et al. 2017). Few studies have identified 100% of the capture-sampled species with metabarcoding assays from a single site or watershed, as discerned here for the Wabash River (A1) and for one of our Maumee River sites (A2). To our knowledge, most other investigations that achieved high resolution likewise employed multiple primer sets and/or were from low-diversity environments (Civade et al. 2016; Shaw et al. 2016; Evans et al. 2017; Fujii et al. 2019).

The trend for higher detection efficiency across broader geographic regions or watersheds compared to findings at individual sampling sites is common (Cilleros et al. 2019; Fujii et al. 2019; Lawson Handley et al. 2019), indicating that overall sampling effort, numbers, volumes, and/or spatial extents with eDNA often are not equivalent to the coverage of many morphological surveys. Shaw et al. (2016) achieved 100% detection regionally for freshwater fish species but had lower success on a per site basis. High identification efficiency (few false negatives and larger number of species uniquely distinguished) in the present results likely was achieved due to our use of multiple metabarcoding assays, since our sampling was limited. The 100% detection efficiency in the Wabash River may have been aided by our collection of duplicate samples; one replicate at each site was collected before and another after the electrofishing transect, which mixed the water and incorporated eDNA from other species.

Sequence identifications have been shown to be related to stringency of bioinformatic filtering (Alberdi et al. 2017; Evans et al. 2017), which were tested here by comparing different error cutoffs for accepting ASVs as “true positives”. Even when all sequences were accepted regardless of frequency, our metabarcoding assays did not identify all species present in the summer-long Maumee River conventional capture-based surveys conducted by the OEPA (A2). Since we analyzed just single 1 L eDNA water samples taken at only four (9% of electrofishing surveys) of those sites, our study demonstrated very high efficiency despite a very minimal sampling effort. Evans et al. (2017) used varying stringencies of bioinformatic filtering based on numbers of samples and/or assays for which species were detected, comparing metabarcoding results to morphological identifications. Their low and moderate stringency bioinformatic methods found all 10 species that were present in capture-based surveys of a small Michigan lake, with three false negatives with their highest stringency criteria. Those authors used rarefaction to show that  $\geq$  eight samples were needed to be processed with all three assays in order for the species accumulation curve to reach an asymptote. Due to sampling constraints, a similar analysis could not be applied here.

Unsurprisingly, more intensive sampling regimes increased the total diversity obtained and improved overlap between metabarcoding assays and capture-based surveys (Evans et al. 2017; Bylemans et al. 2018; Cantera et al. 2019). The time and effort needed to obtain community composition data with the metabarcoding approach here was much less than with traditional methods. On-



site filtering of larger quantities of water also can improve overlap between metabarcoding and morphological identification results (Civade et al. 2016). Some species may be difficult to perceive with eDNA due to their behavior (Barnes and Turner 2016). For example, longnose gar, a lie-in-wait ambush predator, was undetected by our metabarcoding assays in the Maumee River and also in the high diversity aquarium experiments; gar may not shed much eDNA while stationary in the water. As indicated here, the joint use of several targeted and general metabarcoding assays can increase detection when eDNA sampling is limited. Technical replicates (replicate amplifications of the same extractions) also may increase accuracy, however, their effectiveness have been shown to be less than biological replicates (Alberdi et al. 2017).

Traditional capture or visual surveys can be thwarted by physical or environmental conditions (Fujii et al. 2019). The lowest morphological-based species richness value obtained was near the mouth of the Maumee River (Experiment A2: Maumee River 1 at river mile 9.4), in one of the deepest locations sampled, having high suspended solids that decreased visibility (OEPA 2014), under which conditions electrofishing is less effective. In its concomitant eDNA water sample, our assays identified 100% of those species and an additional 20. Fujii et al. (2019) applied metabarcoding assays to backwater lakes in Japan, achieving 100% detection efficiency where capture-based methods were deemed difficult and sampled diversity was low. Different nets and sampling gear possess different biases, selectively capturing some species while leaving others unsampled, with capture avoidances varying with habitat conditions and among species.

The 100% detection efficiency of our metabarcoding assays at MAU 1 may have been due to eDNA samples resolving a larger spatial extent than capture-based methods, particularly in large lotic systems (Cilleros et al. 2019; Fremier et al. 2019). Just three of its 13 species in the morphological sampling surveys were not discerned upstream. Transport of eDNA in large rivers, such as the Maumee River, has been recorded up to 130 km (Pont et al. 2018), which may explain our higher resolution at that site.

Invasive species discovered solely with our metabarcoding assays all were within their known geographic ranges, except for round goby in the Wabash River (A1). The round goby also was identified from eDNA in nearby bait shops (Snyder et al. 2020) and may have expanded its range into the region. Its Wabash River detection occurred in a single assay and sample; thus, there is a possibility that it was a false positive not removed by our pipeline. Other studies also have identified invasive species with metabarcoding assays that were absent from morphological surveys (Zaiko et al. 2018; Fujii et al. 2019).

### Experiment Series A: Community diversity patterns with metabarcoding assays

eDNA metabarcoding assays often have described greater diversity in habitats compared to morphological iden-

tifications from capture-based sampling (see Deiner et al. 2017), as determined here (Experiment Series A). In some cases, species occurring solely in metabarcoding results may result from an incomplete reference database, which can lead to sequence hits for closely related taxa instead of those actually present (Cantera et al. 2019). Our use of a robust *cyt b* database and removal of a few improbable hits for the 12S RNA results circumvented that. Some of the species uniquely found with our metabarcoding assays in the Maumee (A2) or Wabash (A1) rivers either had small body sizes (golden shiner, mooneye) or were benthic (bowfin *Amia calva*, spoonhead sculpin, blue catfish *Ictalurus furcatus*, and round and freshwater tubenose gobies), rendering them less susceptible to electrofishing capture. Several studies likewise have shown that metabarcoding detection of species that often evade traditional capture (Hänfling et al. 2016; Port et al. 2016; Bessey et al. 2020).

Our multi-assay approach and most of the single metabarcoding assays differentiated taxonomic compositions and diversity levels among geographic regions more effectively than did morphological capture-based surveys. This finding is in concert with results from other metabarcoding investigations (Cilleros et al. 2019). The present multi-assay approach also correctly distinguished lentic from lotic habitats, likely due to the greater diversity revealed when results from multiple markers were combined. West et al. (2020) found significant variation in community composition among locations in an extensive metabarcoding survey of a tropical coral reef in the Cocos (Keeling) Islands, Australia, detecting 46 species that previously were undocumented from those sites.

### Experiment Series B: Population genetic patterns from metabarcoding results

Our goby aquarium experiments (Suppl. material 1: Experiment C3) and additional studies have demonstrated the potential of metabarcoding assays to assess population-level diversity (Marshall and Stepien 2019). However, given the false haplotypes present in our goby aquarium experiments, there is a possibility that new diversity in our metabarcoding results (absent from traditional sampling) stemmed from sequence error. Field analyses conducted here suggest that traditional data collection methods may yield different population genetic results than found from metabarcoding assays of eDNA water samples (Experiments B1–2). Although our targeted assay and sampling from the appropriate location in the water column detected all of the haplotypic diversity discerned with traditional single-individual sequencing, the proportions of those haplotypes differed. Factors that can affect species proportions assessed with metabarcoding assays versus physical sampling likewise influenced these population genetic results.

Aquarium eDNA experiments by Tsuji et al. (2018) detected mtDNA control region haplotypes of the ayu *Plecoglossus altivelis*, but despite DADA2's denoising



algorithm (albeit without any frequency based bioinformatic filtering), they discovered 31 false haplotypes, with seven occurring across all 15 replicates. This likely was due to error being non-random on the HTS platform, posing challenges for gathering population genetics data (Nakamura et al. 2011; Schirmer et al. 2016). We also may have some false haplotypes in our eDNA field study findings, meriting further analyses.

To our knowledge, few investigations have examined whether haplotype identities and their frequencies from traditional Sanger sequencing of tissue samples matched those found with metabarcoding assays in the environment. Parsons et al. (2018) compared 88 harbor porpoise *Phocoena phocoena* tissue extractions that had been Sanger sequenced for the mtDNA control region, using a metabarcoding assay of 36 eDNA water samples collected in the fluke prints of diving aggregations. Five of the 28 known haplotypes also were recovered in the metabarcoding results, along with three additional haplotypes (two of which were previously unknown and might constitute false haplotypes). A whale shark *Rhincodon typus* aggregation was sampled offshore in the Arabian Gulf with biopsy spears and sequenced for the complete mtDNA control region by Sigsgaard et al. (2017). Water from eight sites where whale sharks were observed were sampled in triplicate, and the extractions pooled and sequenced using two metabarcoding assays targeting portions of the same gene. One of the assays yielded very similar haplotype frequencies to those determined from tissue sampling, and the other did not, with complete recovery of all haplotypes attributable to the large number of eDNA water samples ( $N = 24$ ) collected.

Haplotypic diversities of populations targeted for metabarcoding assays must be evaluated in order to design primers that are best able to differentiate them. In theory, metabarcoding assays have the ability to distinguish more variation and/or more accurately assess population genetic diversity, due to the much larger numbers of individuals that may be screened, as for thousands of dreissenid mussel larvae by Marshall and Stepien (2019). Accuracy of results is influenced by the assay's design and the targeted taxa, specifically the gene region selected and the match of the primers. Metabarcoding applications for population genetic studies hold promise, but need ground-truthing, careful design, and verification.

## Conclusion

Our metabarcode pipeline revealed high species-level discrimination and low false positive probability employing the multi-assay approach. This was achieved by using multiple primer sets to alleviate false negatives stemming from possible bias, and applying stringent bioinformatic filtering to reduce any false positives from cross-contamination, index-hopping, or sequencing error. Results from these primer sets were combined in a logical way. The multi-assay approach discerned nearly all of the di-

versity sampled over much more extensive traditional electrofishing surveys, and yielded an appreciable number of additional species. We found that our multi- and single assays alike better differentiated among ecological regions and their communities than did morphological identifications from conventional sampling.

Future work likewise should employ a library preparation and bioinformatic pipeline that reduces error using indexing of the initial PCR, and removal of cross-contamination and index-hopping. Such research also should assess effectiveness using positive controls and/or mock communities, for every assay on every run. Multiple values for frequency-based filtering should be evaluated with these positive controls, mock communities, and test samples to determine which performs best. More intensive eDNA water sampling at each location likely would improve the performance of the multi-assay metabarcoding results presented here. Metabarcoding reads potentially could be used as a proxy for proportional taxon abundances within the system, but results are marker dependent and should be interpreted with caution. Current technological limitations may render population genetic analyses using metabarcoding data problematic, but as technology improves, error incidences decline, and longer sequence read lengths become more feasible, this application shows promise.

## Data resources

Scripts for the bioinformatic pipeline and custom BLAST databases are in the Dryad database (<https://doi.org/10.5061/dryad.7m0cfxprx>). FASTQ files for all samples sequenced are in the NCBI Sequence Read Archive (BioProject Accession: PRJNA625378). The Suppl. material 1 contains additional details, including additional experimental results.

## Acknowledgements

This is contribution 4967 from the NOAA Pacific Marine Environmental Laboratory (PMEL) and 2020-1061 from the University of Washington's Joint Institute for the Study of the Atmosphere and Ocean (JISAO). We thank Nathaniel Marshall for aiding in library preparation and pipeline development; Thomas Blomquist, Carson Prichard, and James Willey for early consultation and work on primer development; Edward Roseman for collecting and identifying the species in the larval fish sample; Yuriy Kvach, Matthew Neilson, and Mariusz Sapota for collecting unestablished potential invaders for use in mock communities; Mark Pyron for conducting electrofishing surveys in the Wabash River; David Altfater and the entire Ohio Environmental Protection Agency stream sampling crew for conducting the electrofishing surveys and collecting water samples in the Maumee River; and Keith Wernert for obtaining permission for us to sample display aquarium 2 and providing the census of fishes within. The



laboratory of C.A.S. provided logistical support, especially Shane Yerga-Woolwine and Anna Elz. Additional support for grant and project management was provided by Thomas Ackerman, Frederick Averick, David Butterfield, Frank Calzonetti, Kevin Czajkowski, Timothy Fisher, Darlene Funches, Anna Izzi, Daryl Moorhead, and Scott McBride. Jonathan Bossenbroek, Kerry Naish, and William Von Sigler provided valuable comments on an early version of the manuscript.

The research was funded by USEPA grants GL-00E01149-0 (to C.A.S. and W. Von Sigler) and GL-00E01898 (to C.A.S. and Kevin Czajkowski); the latter was partially subawarded from the University of Toledo to the University of Washington Joint Institute for the Study of the Atmosphere and Ocean (JISAO) for research at the new Genetics and Genomics Group of C.A.S., NOAA Pacific Marine Environmental Laboratory (PMEL). Additional support was provided by NOAA OAR (Oceanic and Atmospheric Research) ‘Omics funding to the Genetics and Genomics Group led by C.A.S. M.R.S. was partially supported by a graduate student fellowship from the University of Toledo (2016–2019), conducted under the advisement of C.A.S., and by JISAO (2019).

## References

- Alberdi A, Aizpurua O, Gilbert MTP, Bohmann K (2017) Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution* 9: 134–147. <https://doi.org/10.1111/2041-210X.12849>
- Allendorf FW, Lundquist LL (2003) Introduction: Population Biology, Evolution, and Control of Invasive Species. *Conservation Biology* 17: 24–30. <https://doi.org/10.1046/j.1523-1739.2003.02365.x>
- Attrill MJ, Depledge MH (1997) Community and population indicators of ecosystem health: targeting links between levels of biological organization. *Aquatic Toxicology* 38:183–197. [https://doi.org/10.1016/S0166-445X\(96\)00839-9](https://doi.org/10.1016/S0166-445X(96)00839-9)
- Balmford A, Gaston KJ (1999) Why biodiversity surveys are good value. *Nature* 398: 204–205. <https://doi.org/10.1038/18339>
- Barnes MA, Turner CR (2016) The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics* 17: 1–17. <https://doi.org/10.1007/s10592-015-0775-4>
- Barnes MA, Turner CR, Jerde CL, Renshaw MA, Chadderton WL, Lodge DM (2014) Environmental conditions influence eDNA persistence in aquatic systems. *Environmental Science and Technology* 48: 1819–1827. <https://doi.org/10.1021/es404734p>
- Begon M, Townsend CR, Harper JL (2006) *Ecology: From Individuals to Ecosystems*, Fourth Edition. Wiley-Blackwell, Malden.
- Bessey C, Jarman SN, Berry O, Olsen YS, Bunce M, Simpson T, Power M, McLaughlin J, Edgar GJ, Keesing J (2020) Maximizing fish detection with eDNA metabarcoding. *Environmental DNA* 00: 1–12. <https://doi.org/10.1002/edn3.74>
- Bylemans J, Gleeson DM, Lintermans M, Hardy CM, Beitzel M, Gilligan DM, Furlan EM (2018) Monitoring riverine fish communities through eDNA metabarcoding: determining optimal sampling strategies along an altitudinal and biodiversity gradient. *Metabarcoding and Metagenomics* 2: e30457. <https://doi.org/10.3897/mbmg.2.30457>
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: high-resolution sample inference from Illumina amplicon data. *Nature Methods* 13: 581–583. <https://doi.org/10.1038/nmeth.3869>
- Cantera I, Cilleros K, Valentini A, Cerdan A, Dejean T, Iribar A, Taberlet P, Vigouroux R, Brosse S (2019) Optimizing environmental DNA sampling effort for fish inventories in tropical streams and rivers. *Scientific Reports* 9: 3085. <https://doi.org/10.1038/s41598-019-39399-5>
- Cilleros K, Valentini A, Allard L, Dejean T, Etienne R, Grenouillet G, Iribar A, Taberlet P, Vigouroux R, Brosse S (2019) Unlocking biodiversity and conservation studies in high-diversity environments using environmental DNA (eDNA): a test with Guianese freshwater fishes. *Molecular Ecology Resources* 19: 27–46. <https://doi.org/10.1111/1755-0998.12900>
- Civade R, Dejean T, Valentini A, Roset N, Raymond J-C, Bonin A, Taberlet P, Pont D (2016) Spatial representativeness of environmental DNA metabarcoding signal for fish biodiversity assessment in a natural freshwater system. *PLoS ONE* 11: e0157366. <https://doi.org/10.1371/journal.pone.0157366>
- Darling JA, Mahon AR (2011) From molecules to management: Adopting DNA-based methods for monitoring biological invasions in aquatic environments. *Environmental Research* 111: 978–988. <https://doi.org/10.1016/j.envres.2011.02.001>
- Deiner K, Altermatt F (2014) Transport distance of invertebrate environmental DNA in a natural river. *PLoS ONE* 9: e88786. <https://doi.org/10.1371/journal.pone.0088786>
- Deiner K, Bik HM, Mächler E, Seymour M, Lacoursière-Roussel A, Altermatt F, Creer S, Bista I, Lodge DM, Vere N de, Pfrender ME, Bernatchez L (2017) Environmental DNA metabarcoding: transforming how we survey animal and plant communities. *ish Aquatic Molecular Ecology* 26: 5872–5895. <https://doi.org/10.1111/mec.14350>
- Dobson AP, Rodriguez JP, Roberts WM, Wilcove DS (1997) Geographic distribution of endangered species in the United States. *Science* 275: 550–553. <https://doi.org/10.1126/science.275.5299.550>
- Elton CS (1966) *The Pattern of Animal Communities*. Springer Netherlands, Rotterdam, The Netherlands.
- Evans NT, Li Y, Renshaw MA, Olds BP, Deiner K, Turner CR, Jerde CL, Lodge DM, Lamberti GA, Pfrender ME (2017) Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. *Canadian Journal of Fisheries and Aquatic Sciences* 74: 1362–1374. <https://doi.org/10.1139/cjfas-2016-0306>
- Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
- Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, Ravel J (2014) An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* 2: 1–6. <https://doi.org/10.1186/2049-2618-2-6>
- Fremier AK, Strickler KM, Parzych J, Powers S, Goldberg CS (2019) Stream transport and retention of environmental DNA pulse releases in relation to hydrogeomorphic scaling factors. *Environmental Science and Technology*. <https://doi.org/10.1021/acs.est.8b06829>
- Fujii K, Doi H, Matsuoka S, Nagano M, Sato H, Yamanaka H (2019) Environmental DNA metabarcoding for fish community analysis in



- backwater lakes: a comparison of capture methods. *PLoS ONE* 14: e0210357. <https://doi.org/10.1371/journal.pone.0210357>
- Gillet B, Cottet M, Destanque T, Kue K, Descoux S, Chanudet V, Hughes S (2018) Direct fishing and eDNA metabarcoding for bio-monitoring during a 3-year survey significantly improves number of fish detected around a South East Asian reservoir. *PLoS ONE* 13: e0208592. <https://doi.org/10.1371/journal.pone.0208592>
- Hänfling B, Handley LL, Read DS, Hahn C, Li J, Nichols P, Blackman RC, Oliver A, Winfield IJ (2016) Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Molecular Ecology* 25: 3101–3119. <https://doi.org/10.1111/mec.13660>
- Hubbs CL, Lagler KF (2007) *Fishes of the Great Lakes Region*, Revised Edition. University of Michigan Press, Ann Arbor, Michigan.
- Jo T, Murakami H, Yamamoto S, Masuda R, Minamoto T (2019) Effect of water temperature and fish biomass on environmental DNA shedding, degradation, and size distribution. *Ecology and Evolution* 9: 1135–1146. <https://doi.org/10.1002/ece3.4802>
- Kelly RP, Closek CJ, O'Donnell JL, Kralj JE, Shelton AO, Samhour JF (2017) Genetic and manual survey methods yield different and complementary views of an ecosystem. *Frontiers in Marine Science* 3: 1–11. <https://doi.org/10.3389/fmars.2016.00283>
- Khodakov D, Wang C, Zhang DY (2016) Diagnostics based on nucleic acid sequence variant profiling: PCR, hybridization, and NGS approaches. *Advances in Drug Delivery Reviews* 105: 3–19. <https://doi.org/10.1016/j.addr.2016.04.005>
- Klymus KE, Marshall NT, Stepien CA (2017) Environmental DNA (eDNA) metabarcoding assays to detect invasive invertebrate species in the Great Lakes. *PLoS ONE* 12: e0177643. <https://doi.org/10.1371/journal.pone.0177643>
- Kolar C, Chapman D, Courtenay W, Housel C, Williams J, Jennings D (2005) Asian carps of the genus *Hypophthalmichthys* (Pisces, Cyprinidae) — a biological synopsis and environmental risk assessment. National Invasive Species Council Materials. <http://digital-commons.unl.edu/natinvasive/5>
- Lawson Handley L, Read DS, Winfield IJ, Kimbell H, Johnson H, Li J, Hahn C, Blackman R, Wilcox R, Donnelly R, Szitenberg A, Hänfling B (2019) Temporal and spatial variation in distribution of fish environmental DNA in England's largest lake. *Environmental DNA* 1: 26–39. <https://doi.org/10.1002/edn3.5>
- MacConaill LE, Burns RT, Nag A, Coleman HA, Slevin MK, Giorda K, Light M, Lai K, Jarosz M, McNeill MS, Ducar MD, Meyerson M, Thorner AR (2018) Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. *BMC Genomics* 19: 1–30. <https://doi.org/10.1186/s12864-017-4428-5>
- Margules CR, Pressey RL, Williams PH (2002) Representing biodiversity: Data and procedures for identifying priority areas for conservation. *Journal of Biosciences* 27: 309–326. <https://doi.org/10.1007/BF02704962>
- Marshall NT, Stepien CA (2019) Invasion genetics from eDNA and thousands of larvae: A targeted metabarcoding assay that distinguishes species and population variation of zebra and quagga mussels. *Ecology and Evolution* 9: 1–24. <https://doi.org/10.1002/ece3.4985>
- Miya M, Sato Y, Fukunaga T, Sado T, Poulsen JY, Sato K, Minamoto T, Yamamoto S, Yamanaka H, Araki H, Kondoh M, Iwasaki W (2015) MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *Royal Society Open Science* 2: 150088. <https://doi.org/10.1098/rsos.150088>
- Morin PJ (2009) *Community Ecology*. John Wiley & Sons, Oxford.
- Müllner D (2013) Fastcluster: fast hierarchical, agglomerative clustering routines for R and Python. *Journal of Statistical Software* 53: 1–18. <https://doi.org/10.18637/jss.v053.i09>
- Myers N, Mittermeier RA, Mittermeier CG, Fonseca GAB da, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* 403: 1–853. <https://doi.org/10.1038/35002501>
- Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, Shiwa Y, Ishikawa S, Linak MC, Hirai A, Takahashi H, Altaf-Ul-Amin M, Ogasawara N, Kanaya S (2011) Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res* 39: e90–e90. <https://doi.org/10.1093/nar/gkr344>
- NOAA (2019) GLANSIS – Great Lakes Aquatic Nonindigenous Species Information System. <https://www.glerl.noaa.gov/glansis/> [Accessed 31 July 2020]
- OEPA (2014) Biological and Water Quality Study of the Maumee River and Auglaize River 2012–2013. Ohio EPA Division of Surface Water. [https://www.epa.ohio.gov/Portals/35/documents/MaumeeTSD\\_2014.pdf](https://www.epa.ohio.gov/Portals/35/documents/MaumeeTSD_2014.pdf)
- OEPA (2015) Biological Criteria for the Protection of Aquatic Life: Volume III. Standardized Biological Field Sampling and Laboratory Methods for Assessing Fish and Macroinvertebrate Communities. In: Ohio EPA Division of Surface Water. [https://epa.ohio.gov/portals/35/documents/BioCrit15\\_Vol3.pdf#page41](https://epa.ohio.gov/portals/35/documents/BioCrit15_Vol3.pdf#page41)
- Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2019) Package 'vegan.' CRAN. <https://cran.r-project.org/web/packages/vegan> [Accessed 31 July 2020]
- Parsons KM, Everett M, Dahlheim M, Park L (2018) Water, water everywhere: environmental DNA can unlock population structure in elusive marine species. *Royal Society of Open Science* 5: 180537. <https://doi.org/10.1098/rsos.180537>
- Pont D, Rocle M, Valentini A, Civade R, Jean P, Maire A, Roset N, Schabuss M, Zornig H, Dejean T (2018) Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. *Scientific Reports* 8: 10361. <https://doi.org/10.1038/s41598-018-28424-8>
- Port JA, O'Donnell JL, Romero-Maraccini OC, Leary PR, Litvin SY, Nickols KJ, Yamahara KM, Kelly RP (2016) Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular Ecology* 25: 527–541. <https://doi.org/10.1111/mec.13481>
- R Core Team (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.
- Rice WR (1989) Analyzing Tables of Statistical Tests. *Evolution* 43: 223–225. <https://doi.org/10.2307/2409177>
- Robin ED, Wong R (1988) Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *Journal of Cellular Physiology* 136: 507–513. <https://doi.org/10.1002/jcp.1041360316>
- Schirmer M, D'Amore R, Ijaz UZ, Hall N, Quince C (2016) Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinformatics* 17: 1–125. <https://doi.org/10.1186/s12859-016-0976-y>
- Shaw JLA, Clarke LJ, Wedderburn SD, Barnes TC, Weyrich LS, Cooper A (2016) Comparison of environmental DNA metabar-



- coding and conventional fish survey methods in a river system. *Biological Conservation* 197: 131–138. <https://doi.org/10.1016/j.biocon.2016.03.010>
- Sigsgaard EE, Nielsen IB, Bach SS, Lorenzen ED, Robinson DP, Knudsen SW, Pedersen MW, Jaidah MA, Orlando L, Willerslev E, Møller PR, Thomsen PF (2017) Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nature Ecology & Evolution* 1: 0004. <https://doi.org/10.1038/s41559-016-0004>
- Simon T (2006) Biodiversity of fishes in the Wabash River: Status, indicators, and threats. *Proceedings of the Indiana Academy of Science* 115: 136–148.
- Smart AS, Weeks AR, Rooyen AR van, Moore A, McCarthy MA, Tingley R (2016) Assessing the cost-efficiency of environmental DNA sampling. *Methods in Ecology and Evolution* 7: 1291–1298. <https://doi.org/10.1111/2041-210X.12598>
- Snyder MR, Stepien CA (2017) Genetic patterns across an invasion's history: a test of change versus stasis for the Eurasian round goby in North America. *Molecular Ecology* 26: 1075–1090. <https://doi.org/10.1111/mec.13997>
- Snyder MR, Stepien CA, Marshall NT, Scheppler HB, Black CL, Czajkowski KP (2020) Detecting aquatic invasive species in bait and pond stores with targeted environmental (e) DNA high-throughput sequencing metabarcoding assays: angler, retailer, and manager implications. *Biological Conservation* 243: 108430. <https://doi.org/10.1016/j.biocon.2020.108430>
- Stepien CA, Snyder MR, Elz AE (2019) Invasion genetics of the silver carp *Hypophthalmichthys molitrix* across North America: Differentiation of fronts, introgression, and eDNA metabarcoding detection. *PLoS ONE* 14: e0203012. <https://doi.org/10.1371/journal.pone.0203012>
- Stoeckle MY, Soboleva L, Charlop-Powers Z (2017) Aquatic environmental DNA detects seasonal fish abundance and habitat preference in an urban estuary. *PLoS ONE* 12: e0175186. <https://doi.org/10.1371/journal.pone.0175186>
- Suzuki R, Shimodaira H (2015) Package “pvclust”: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling. CRAN. <https://cran.r-project.org/web/packages/pvclust/pvclust.pdf> [Accessed 31 July 2020]
- Thomsen PF, Møller PR, Sigsgaard EE, Knudsen SW, Jørgensen OA, Willerslev E (2016) Environmental DNA from seawater samples correlate with trawl catches of subarctic, deepwater fishes. *PLoS ONE* 11: e0165252. <https://doi.org/10.1371/journal.pone.0165252>
- Tsuji S, Miya M, Ushio M, Sato H, Minamoto T, Yamanaka H (2018) Evaluating intraspecific diversity of a fish population using environmental DNA: An approach to distinguish true haplotypes from erroneous sequences. *bioRxiv* 429993. <https://doi.org/10.1101/429993>
- Turner CR, Barnes MA, Xu CCY, Jones SE, Jerde CL, Lodge DM (2014) Particle size distribution and optimal capture of aqueous microbial eDNA. *Methods in Ecology and Evolution* 5: 676–684. <https://doi.org/10.1111/2041-210X.12206>
- Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, Bellemain E, Besnard A, Coissac E, Boyer F, Gaboriaud C, Jean P, Poulet N, Roset N, Copp GH, Geniez P, Pont D, Argillier C, Baudoin J-M, Peroux T, Crivelli AJ, Olivier A, Acqueberge M, Brun ML, Møller PR, Willerslev E, Dejean T (2016) Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology* 25: 929–942. <https://doi.org/10.1111/mec.13428>
- van Bleijswijk JDL, Engelmann JC, Klunder L, Witte HJ, Witte JI, Veer HW van der (2020) Analysis of a coastal North Sea fish community: Comparison of aquatic environmental DNA concentrations to fish catches. *Environmental DNA* 00: 1–17. <https://doi.org/10.1002/edn3.67>
- van Bochove K, Bakker FT, Beentjes KK, Hemerik L, Vos RA, Gravendeel B (2020) Organic matter reduces the amount of detectable environmental DNA in freshwater. *Ecology and Evolution* 10: 3647–3654. <https://doi.org/10.1002/ece3.6123>
- West KM, Stat M, Harvey ES, Skepper CL, DiBattista JD, Richards ZT, Travers MJ, Newman SJ, Bunce M (2020) eDNA metabarcoding survey reveals fine-scale coral reef community variation across a remote, tropical island ecosystem. *Molecular Ecology* 29: 1069–1086. <https://doi.org/10.1111/mec.15382>
- Williams KE, Huyvaert KP, Piaggio AJ (2017) Clearing muddied waters: Capture of environmental DNA from turbid waters. *PLoS ONE* 12: e0179282. <https://doi.org/10.1371/journal.pone.0179282>
- Wu L, Wen C, Qin Y, Yin H, Tu Q, Nostrand JDV, Yuan T, Yuan M, Deng Y, Zhou J (2015) Phasing amplicon sequencing on Illumina Miseq for robust environmental microbial community analysis. *BMC Microbiology* 15: 1–125. <https://doi.org/10.1186/s12866-015-0450-4>
- Xiong W, Li H, Zhan A (2016) Early detection of invasive species in marine ecosystems using high-throughput sequencing: technical challenges and possible solutions. *Marine Biology* 163: 1–139. <https://doi.org/10.1007/s00227-016-2911-1>
- Zaiko A, Pochon X, Garcia-Vazquez E, Olenin S, Wood SA (2018) Advantages and limitations of environmental DNA/RNA tools for marine biosecurity: management and surveillance of non-indigenous species. *Frontiers in Marine Science* 5: 1–17. <https://doi.org/10.3389/fmars.2018.00322>
- Zar JH (2010) *Biostatistical Analysis*, Fifth Edition. Pearson, Upper Saddle River, NJ.
- Zinger L, Bonin A, Alsos IG, Bálint M, Bik H, Boyer F, Chariton AA, Creer S, Coissac E, Deagle BE, Barba MD, Dickie IA, Dumbrell AJ, Ficetola GF, Fierer N, Fumagalli L, Gilbert MTP, Jarman S, Järpönen A, Kausarud H, Orlando L, Pansu J, Pawłowski J, Tedersoo L, Thomsen PF, Willerslev E, Taberlet P (2019) DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology* 28: 1857–1862. <https://doi.org/10.1111/mec.15060>

### Supplementary material 1

#### Additional methods, results, figures, and tables

Authors: Matthew R. Snyder, Carol A. Stepien

Data type: Additional methods, results, figures, tables

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited. Link: <https://doi.org/10.3897/mbmg.4.53455.suppl1>